

## **8.3 列联表与独立性检验**

### **8. 3.1 分类变量与列联表**

课标要求	素养要求
<p>1.通过实例，理解<math>2\times 2</math>列联表的统计意义.</p> <p>2.理解判断两个分类变量是否有关系的常用方法.</p>	<p>通过学习<math>2\times 2</math>列联表，提升数学抽象、直观想象及数据分析素养.</p>

## 新知探究

► 情境引入

饮用水的质量是人类普遍关心的问题，根据统计，饮用优质水的518人中，身体状况优秀的有466人，饮用一般水的312人中，身体状况优秀的有218人.



**问题** 人的身体健康状况与饮用水的质量之间有关系吗？

**提示** 我们可以根据 $2 \times 2$ 列联表找到人的身体健康与饮用水之间的关系，也就是本节课所要学习的内容.

## ► 知识梳理

### 1. 分类变量

这里所说的变量和值不一定是具体的数值，例如：性别变量，其取值为男和女两种

我们经常会使用一种特殊的随机变量，以区别不同的现象或性质，这类随机变量称为分类变量，分类变量的取值可以用实数表示。

## 2. $2 \times 2$ 列联表

在实践中，由于保存原始数据的成本较高，人们经常按研究问题的需要，将数据分类统计，并做成表格加以保存，我们将这类数据统计表称为 $2 \times 2$ 列联表， $2 \times 2$ 列联表给出了成对分类变量数据的交叉分类频数。

一般地，假设有两个分类变量 $X$ 和 $Y$ ，它们的取值分别为 $\{x_1, x_2\}$ 和 $\{y_1, y_2\}$ ，其 $2 \times 2$ 列联表为

	$y_1$	$y_2$	合计
$x_1$	$a$	$b$	$a + b$
$x_2$	$c$	$d$	$c + d$
合计	$a + c$	$b + d$	$a + b + c + d$

### 3.等高堆积条形图

等高条形图和表格相比，更能直观地反映出两个分类变量间是否相互影响，常用等高条形图展示列联表数据的频率特征，依据频率稳定于概率的原理，我们可以推断结果。

## 拓展深化

### [微判断]

1. 分类变量中的变量与函数中的变量是同一概念. ( × )

**提示** 分类变量中的变量是指一定范围内的两种现象或性质，与函数中的变量不是同一概念。

2. 列联表中的数据是两个分类变量的频数. ( ✓ )

3. 列联表、频率分析法、等高条形图都可初步分析两分类变量是否有关系.

( ✓ )

[微训练]

1. 下列不是分类变量的是( )

- A. 近视                  B. 成绩  
C. 血压                  D. 饮酒

**解析** 近视变量有近视与不近视两种类别，血压变量有异常、正常两种类别，饮酒变量有饮酒与不饮酒两种类别。故选B.

**答案** B

2. 某校为了检验高中数学新课程改革的成果，在两个班进行教学方式的对比试验，两个月后进行了一次检测，试验班与对照班成绩统计如 $2 \times 2$ 列联表所示(单位：人)，则其中 $m =$ \_\_\_\_\_， $n =$ \_\_\_\_\_.

	80分及80分以上	80分以下	合计
试验班	32	18	50
对照班	24	$m$	50
合计	56	44	$n$

解析 由题意得  $\begin{cases} 24+m=50, \\ 56+44=n, \end{cases}$

解得  $\begin{cases} m=26, \\ n=100. \end{cases}$

答案 26 100

[微思考]

1. 是否吸烟、是否患肺癌是什么变量？

提示 分类变量.

2. 吸烟与患肺癌之间的关系还是前面我们研究的线性相关关系吗？

提示 不是.

题型一 用  $2 \times 2$  列联表分析两分类变量间的关系

**【例 1】** 在对人们饮食习惯的一次调查中，共调查了 124 人，其中六十岁以上的 70 人，六十岁以下的 54 人．六十岁以上的人中有 43 人的饮食以蔬菜为主，另外 27 人则以肉类为主；六十岁以下的人中有 21 人饮食以蔬菜为主，另外 33 人则以肉类为主．请根据以上数据作出饮食习惯与年龄的列联表，并利用  $\frac{a}{a+b}$  与  $\frac{c}{c+d}$  判断二者是否有关系．

解  $2 \times 2$ 列联表如下:

	年龄在六十岁以上	年龄在六十岁以下	合计
饮食以蔬菜为主	43	21	64
饮食以肉类为主	27	33	60
合计	70	54	124

将表中数据代入公式得

$$\frac{a}{a+b} = \frac{43}{64} = 0.671\ 875, \frac{c}{c+d} = \frac{27}{60} = 0.45.$$

显然二者数据具有较为明显的差距，据此可以在某种程度上认为饮食习惯与年龄有关系。

**规律方法** (1)作  $2 \times 2$  列联表时, 关键是对涉及的变量分清类别. 计算时要准确无误.

(2)利用  $2 \times 2$  列联表分析两个分类变量间的关系时, 首先要根据题中数据获得  $2 \times 2$  列联

表, 然后根据频率特征, 即将  $\frac{a}{a+b}$  与  $\frac{c}{c+d}$   $\left( \frac{b}{a+b}$  与  $\frac{d}{c+d} \right)$  的值相比, 直观地反映出两个分

类变量间是否相互影响, 但方法较粗劣.

【训练1】 假设有两个分类变量 $X$ 与 $Y$ ，它们的可能取值分别为 $\{x_1, x_2\}$ 和 $\{y_1, y_2\}$ ，其 $2 \times 2$ 列联表为：

	$y_1$	$y_2$
$x_1$	10	18
$x_2$	$m$	26

则当 $m$ 取下面何值时， $X$ 与 $Y$ 的关系最弱( )

- A. 8                      B. 9                      C. 14                      D. 19

解析 由 $10 \times 26 \approx 18m$ ，解得 $m \approx 14.4$ ，所以当 $m = 14$ 时， $X$ 与 $Y$ 的关系最弱。

答案 C

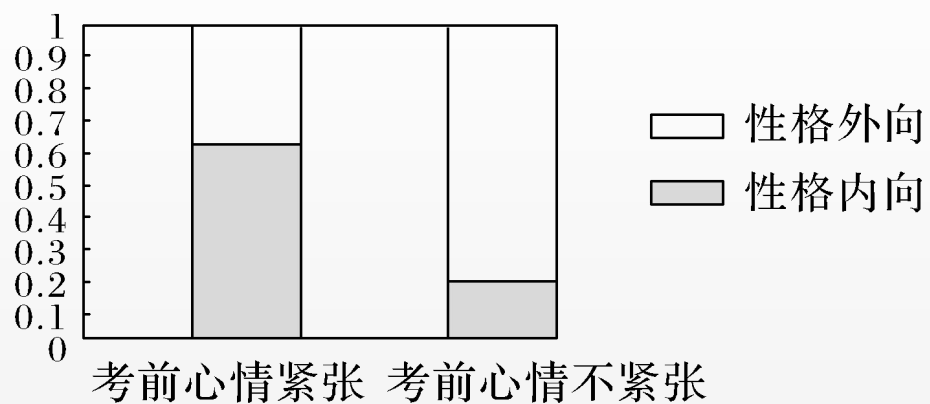
## 题型二 用等高堆积条形图分析两分类变量间的关系

**【例2】** 某学校对高三学生作了一项调查发现：在平时的模拟考试中，性格内向的学生426人中有332人在考前心情紧张，性格外向的学生594人中有213人在考前心情紧张，作出等高条形图，利用图形判断考前心情紧张与性格类型是否有关系。

**解** 作列联表如下：

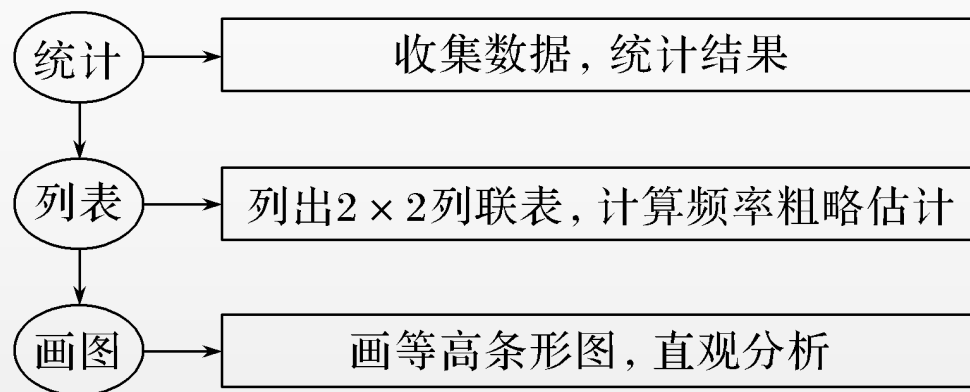
	性格内向	性格外向	合计
考前心情紧张	332	213	545
考前心情不紧张	94	381	475
合计	426	594	1 020

相应的等高堆积条形图如图所示：



图中阴影部分表示考前心情紧张与考前心情不紧张中性格内向的人数的比例，从图中可以看出考前心情紧张的样本中性格内向的人数占的比例比考前心情不紧张样本中性格内向的人数占的比例高，可以认为考前心情紧张与性格类型有关。

规律方法 利用等高堆积条形图判断两个分类变量是否相关的步骤:

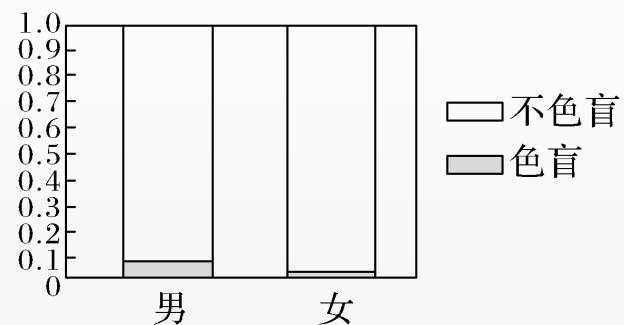


**【训练2】** 在调查的480名男人中有38人患色盲，520名女人中有6名患色盲，试利用图形来判断色盲与性别是否有关？

**解** 根据题目给出的数据作出如下的列联表：

	色盲	不色盲	合计
男	38	442	480
女	6	514	520
合计	44	956	1 000

根据列联表作出相应的等高堆积条形图：



从等高堆积条形图来看，在男人中患色盲的比例要比在女人中患色盲的比例大得多，因此，我们认为患色盲与性别是有关系的。

## 一、素养落地

1. 通过本节课的学习，进一步提升数学抽象、直观想象及数据分析素养.
2. 列联表与等高堆积条形图

列联表由两个分类变量之间频率大小的差异说明这两个变量之间是否有相关关系，而利用等高堆积条形图能形象直观地反映它们之间的差异，进而推断它们之间是否具有相关关系.

## 二、素养训练

1. 与表格相比, 能更直观地反映出相关数据总体状况的是( )

A. 列联表

B. 散点图

C. 残差图

D. 等高堆积条形图

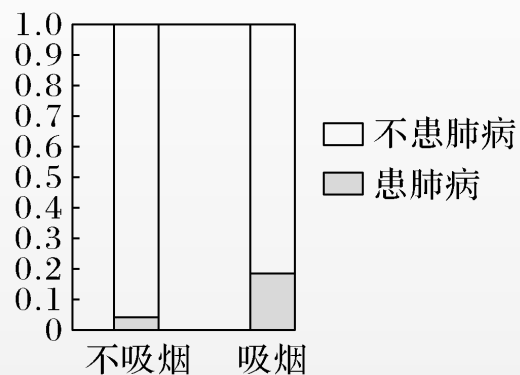
答案 D

2. 在一项有关医疗保健的社会调查中，发现调查的男性为530人，女性为670人，其中男性中喜欢吃甜食的为117人，女性中喜欢吃甜食的为492人，则性别与喜欢吃甜食的 $2 \times 2$ 列联表为\_\_\_\_\_.

**答案**

	喜欢吃甜食	不喜欢吃甜食	合计
男	117	413	530
女	492	178	670
合计	609	591	1 200

3.根据如图所示的等高堆积条形图可知吸烟与患肺病\_\_\_\_\_关系(填“有”或“没有”).



**解析** 从等高条形图上可以明显地看出吸烟患肺病的频率远远大于不吸烟患肺病的频率.

**答案** 有

4. (多选题)下面是一个 $2 \times 2$ 列联表:

	$y_1$	$y_2$	合计
$x_1$	$a$	21	73
$x_2$	2	25	27
合计	$b$	46	100

则表中 $a = \underline{\hspace{2cm}}$ ,  $b = \underline{\hspace{2cm}}$ .

解析 由题意得  $\begin{cases} a+21=73, \\ a+2=b, \end{cases}$  解得  $\begin{cases} a=52, \\ b=54. \end{cases}$

答案 52 54

5. 为考察某种药物预防疾病的效果进行动物试验，得到如下列联表：

	患病	未患病	合计
服用药	10	45	55
未服用药	20	30	50
合计	30	75	105

试用等高条形图分析服用药和患病之间是否有关系。

解 根据列联表所给的数据可得出服用药患病的频率为 $\frac{10}{55} \approx 0.18$ ，未服用药患病的频率为 $\frac{20}{50} = 0.4$ ，两者的差距是 $|0.18 - 0.4| = 0.22$ ，两者相差很大，作出等高条形图如图所示，因此服用药与患病有关系。

